TP 1

On rappelle que la commande `help` X permet d'obtenir une aide détaillée sur la commande X.

Exercice 1 Visualisation des lois discrètes

1. (a) Exécuter le code suivant, qui permet de visualiser les résultats de 100 tirages indépendants selon la loi binomiale $B(20, 1/5)$ et de comparer avec la valeur théorique.

```
X = grand(100,1,"bin",20,0.2);
F = tabul(X)   // tableau des effectifs
F(:,2) = F(:,2)/100
plot2d3(F(:,1),F(:,2))
plot2d(0:20,binomial(0.2,20),style=[-1])
```

(b) Qu'observe-t-on quand on remplace 100 par une valeur plus grande ? Quel théorème justifie cela ?

Correction : Les fréquences s'approchent de la distribution théorique. C'est la loi des grands nombres : si $X_1, \ldots, X_n$ sont iid de loi $B(20, 1/5)$,

$$\mathbf{P}(X = k) = \lim \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{X_i=k}.$$

(c) Visualisez de manière similaire les loi de Poisson $P(1)$, $P(4)$ et $P(50)$.

Correction :

```
p=1 //paramètre de la loi de Poisson
n=100 //nombre de tirages
X=grand(n,1,"poi",p);
F=tabul(X);
m=min(F(:,1));
M=max(F(:,1));
F(:,2)=F(:,2)/n;
plot2d3(F(:,1),F(:,2))
Y=exp(-p)*(p.^(m:M))./factorial(m:M);
plot2d(m:M,Y,style=[-1])
```

(d) A quoi ressemble le graphique obtenu pour $P(4)$? Pourquoi ?

Correction : Il ressemble à celui obtenu à la question 1, parce que la loi $B(n, p/n)$ converge en loi vers la loi $P(p)$ (et $4 = 20 * .2$).

(e) A quoi ressemble le graphique obtenu pour $P(50)$? (*) Pourquoi ?

Correction : Il ressemble à une gaussienne centrée autour de 50. En effet, si on suppose que $k = p + \ell$ avec $\ell \ll p$, en utilisant la formule de Stirling et en faisant un développement limité, on trouve

$$e^{-p} p^k / k! \sim e^{-p} \frac{p^k}{\sqrt{2\pi p} e^{-p-\ell} p^k (1 + \ell/p)^k} \sim \frac{e^{-\ell^2/p}}{\sqrt{2\pi p}}.$$

(f) On a vu le résultat suivant : si X et Y sont indépendantes, avec $X \sim P(\lambda)$ et $Y \sim P(\mu)$, alors $X + Y \sim P(\lambda + \mu)$. Vérifiez visuellement ce résultat pour $\lambda = 1$, $\mu = 2$.

Correction :

```
n=1000 //nombre d'observations
X=grand(n,1,"poi",1);
Y=grand(n,1,"poi",2);
Z=X+Y;
F=tabul(Z);
```

```
m=min(F(:,1));
M=max(F(:,1));
F(:,2)=F(:,2)/n;
plot2d3(F(:,1),F(:,2))
Zth=exp(-3)*(3.^(m:M))./factorial(m:M);
plot2d(m:M,Zth,style=[-1])
```

(g) (*) Comparer le temps mis par scilab pour générer des v.a. de loi $B(10^2, 10^{-2})$ et $B(10^9, 10^{-9})$. Pourquoi le résultat est-il surprenant ?

Correction : C'est sensiblement le même temps. On peut le vérifier en utilisant les commandes `tic` *et* `toc`*.*

```
tic
X=grand(10^6,1,"bin",100,.01);
toc
```

Puis même commande avec les autres paramètres. Si la simulation de $B(n, p)$ se fait en simulant n v.a. indépendantes de loi $\mathrm{Ber}(p)$ *(ce qui est le plus naturel), il devrait falloir 10^7 fois plus de temps pour simuler $B(10^9, 10^{-9})$ que $B(100, .01)$.*

2. (a) En utilisant la commande `cumsum`, modifier le code de la question (1a) pour qu'il trace la fonction de répartition (empirique et théorique). *Correction :*

```
n=1000; //nombre de tirages
X = grand(n,1,"bin",20,0.2);
F = tabul(X,"i")  ; // tableau des effectifs rangés du plus petit au plus grand
F(:,2)=cumsum(F(:,2));
F(:,2)=F(:,2)/n;
plot2d3(F(:,1),F(:,2))
plot2d(0:20,cumsum(binomial(0.2,20)),style=[-1])
```

(b) Si $X \sim P(2)$ et $Y \sim P(3)$, tracer sur un même graphe les fonctions de répartitions F_X et F_Y et conjecturer une inégalité entre ces fonctions. (*) Sauriez-vous la démontrer ? *Correction :*

```
FX=cumsum(exp(-2)*(2.^(0:20))./factorial(0:20));
FY=cumsum(exp(-3)*(3.^(0:20))./factorial(0:20));
plot2d((0:20),[F2;F3]', style=[-1,-2])
```

On observe $F_X \geqslant F_Y$. Le plus simple pour le montrer est de se rappeler qu'on peut réaliser Y comme $X + Z$, où $Z \sim P(1)$ est une v.a. indépendante de X. Avec ce couplage, il est clair que $Y \geqslant X$, et donc $\mathbf{P}(Y \leqslant t) \leqslant \mathbf{P}(X \leqslant t)$. Ce n'est pas évident de faire un calcul direct.

Exercice 2 Loi des grands nombres

Le but de cet exercice est de visualiser la loi forte des grands nombres. On se donne une suite de v.a. i.i.d. $(X_n)_{n\geqslant 1}$, et on s'intéresse au comportement asymptotique de

$$Y_n := \frac{X_1 + \cdots + X_n}{n}$$

en fonction de différents choix pour la loi de X_1

1. (a) On suppose que X_n suit la loi uniforme dans l'intervalle $[0, 2]$. Utiliser la commande `grand` pour générer un tableau `X` contenant les valeurs des v.a. $(X_n)_{1\leqslant n\leqslant N}$, pour $N = 10^5$.

Correction :

```
N=10^5;
X=grand(N,1,"unf",0,2);
```

(b) En une seule instruction, générez un tableau `Y` contenant les valeurs des v.a. $(Y_n)_{1\leqslant n\leqslant N}$.

Correction :

```
Y=cumsum(X)./(1:N)';
```

(c) Visualisez le comportement de la suite (Y_n) à l'aide de la commande `plot([1:N],Y)`. Qu'observe-t-on ?

Correction : Après quelques oscillations, la suite se stabilise autour de $1 = \mathbf{E}[U]$. Ça tombe bien, c'est ce que prédit la loi des grands nombres.

2. Reprendre les questions précédentes lorsque X_n suit la loi de $1/U$, où U suit la loi uniforme dans l'intervalle $[0,2]$. Qu'observe-t-on ? (*) Sauriez-vous le démontrer ?

   Correction :

```
N=10^5;
X=grand(N,1,"unf",0,2);
X=1 ./X;
Y=cumsum(X)./(1:N)';
plot([1:N],Y')
```

   La suite n'a pas l'air de se stabiliser, mais plutôt de faire de temps en temps des sauts vers le haut entre lesquels elle décroît. Remarquons que $\mathbf{E}[1/U] = \int_0^1 \frac{1}{x}dx = +\infty$*, donc la loi des grands nombres ne s'applique pas. On a vu en fait (exo 6 TD47) que dans ce cas* $\limsup Y_n = +\infty$ *presque sûrement.*

3. Reprendre les questions précédentes lorsque X_n suit la loi de $1/V$, où V suit la loi uniforme dans l'intervalle $[-1,1]$. Qu'observe-t-on ?

   Correction :

```
N=10^5;
X=grand(N,1,"unf",-1,1);
X=1 ./X;
Y=cumsum(X)./(1:N)';
plot([1:N],Y')
```

La suite a un comportement similaire à l'exemple précédent, mais les sauts sont aussi bien positifs que négatifs. On peut remarquer qu'on est dans un cas où $\mathbf{E}[1/U] = 0$ *car la densité est symétrique, mais* $1/U$ *n'est pas dans* L^1*.*

Exercice 3 Théorème central limite

1. Reprendre le code de l'exercice précédent pour visualiser la suite

$$Z_n := \frac{X_1 + \cdots + X_n - n}{\sqrt{n}},$$

   où les (X_n) sont i.i.d. de loi uniforme sur $[0,2]$. Qu'observe-t-on ?

   Correction :

```
N=10^5;
X=2*grand(N,1,"unf",0,2);
Y=(cumsum(X)-(1:N)')./sqrt((1:N)');
plot([1:N],Y')
```

2. On suppose désormais que les (X_i) sont i.i.d. de loi uniforme sur $\{-1,1\}$, et on pose $S_N = \frac{X_1+...+X_N}{\sqrt{N}}$. Comprendre et exécuter le code suivant qui permet de tracer l'histrogramme de n copies indépendantes de la variable Z_N. Faites varier n et N ; qu'observe-t-on ?

```
n=10000;
N=100;
X=(2*grand(1,n,'bin',N,1/2)-N)/sqrt(N);
histplot(int(sqrt(n)),X)
```

   Correction : Pour $N = 100$*, l'histogramme a plein de trous. C'est lié au fait que les variables* X_n *sont discrètes, et* 100 *n'est pas un nombre assez grand pour approcher le continu à l'échelle* $1/\sqrt{10000}$*. Quand on augmente* N*, les trous disparaissent avec toujours certaines valeurs privilégiées (en grands pics). Pour* $N = 10^6$*, la structure discrète a disparu. Quand au contraire* n *diminue, l'histogramme devient de plus en plus grossier.*

3. Que fait le code suivant ? Pourquoi est-ce pertinent ?

```
T=[-3:0.1:3];
U=exp(-T.^2/2)/sqrt(2*%pi);
plot(T,U)
```

Correction : Le code trace une discrétisation de la densité de la loi normale centrée réduite. D'après le TCL, si N est suffisamment grand, la loi de Z_N doit être proche de $\mathcal{N}(0,1)$. Avec un nombre d'observations n suffisamment grand, l'histogramme tracé à la question précédente doit donc bien approximer la courbe tracée ici.

Exercice 4 Le paradoxe des anniversaires

La probabilité pour que, dans un groupe de K personnes, deux au moins aient la même date d'anniversaire est beaucoup plus élevée que ce que l'on pourrait croire. On modélise les K dates d'anniversaire par des variables aléatoires i.i.d. de loi uniforme dans $\{1,\ldots,365\}$ (on ne tient pas compte des années bissextiles).

1. Ecrire la function `p=pcoincide(K)` renvoyant (par un calcul exact !) cette probabilité p de coïncidence en fonction du nombre K de gens dans le groupe.

   Correction :

```
function [p]=pcoincide(k)
if k>365 then p=1;
// Si plus de 365 personnes, c'est clair qu'il y a forcément coïncidence.
elseif k==1 then p=0;
// Avec une personne la probabilité de coincidence est 0.
else  p=1-(1-pcoincide(k-1))*((366-k)/365);
// A chaque étape supplémentaire, la probabilité de "non-coincidence" est
// multipliéé par (366-k)/365. Concrètement: tant qu'il n'y a pas coïncidence,
// il y a k dates occupées, donc la probabilité de tomber sur une date
// inoccupée est à chaque étape multipliée par (366-k)/365.)
end
endfunction
```

2. Tracer la courbe de ces probabilités pour un nombre d'individus compris entre 2 et 365 avec la commande `plot2d(2:365, feval(2:365, pcoincide))`. A partir de combien dépasse-t-elle 50%, 75%... ?

   Correction : 50% dépassés à 23 et 75 % à 32.

3. Pour vérifier statistiquement la valeur de `pcoincide(23)`, on simulera un grand nombre N (par exemple 1000) de populations de 23 personnes, dont les dates de naissance seront tirées au hasard avec probabilité uniforme sur les 365 jours de l'année, on calculera la fréquence de réalisation de l'événement "coïncidence" et on constatera qu'elle est proche de la valeur théorique.

   Correction :

```
function [q]=coincidence(N)
counter=0;
// Le counter correspond au nombre d'expériences avec coïncidence
for i=1:N
    X=grand(1,23,"uin",1,365);
    // On fait le tirage au sort
    X=gsort(X)
    // On range notre vecteur aléatoire
       for k=1:22
       // S'il y a deux dates consécutives qui coïncident, on rajoute 1
       // au compteur et on passe à la prochaine expérience.
       if X(k)==X(k+1) then counter=counter+1, break;
       end
    end
end
q=counter/N
// q est la proportion des nombres d'expérience avec coïncidence.
endfunction
```

Exercice 5 Collection de vignettes

Une marque de céréales décide d'inclure dans chaque paquet une image de l'un des N mathématiciens les plus célèbres pour inciter les enfants à les collectionner. On suppose que les images sont placées dans les paquets de manière indépendante selon la loi uniforme sur $\{1,\ldots,N\}$. Ecrire une fonction `vignette(N)` qui renvoie une simulation du nombre X_N de paquets qu'il a fallu acheter pour avoir la collection complète des N vignettes. Saurez-vous deviner à quelle vitesse X_N tend vers l'infini ?

Correction : La v.a. X_N s'écrit comme

$$X_N = \sum_{i=1}^{N} T_i$$

où T_i est le nombre d'achats nécessaires pour obtenir la i ième vignette sachant qu'on en a $i-1$. Les v.a. $(T_i)_{1\leqslant i\leqslant N}$ suivent une loi géométrique de paramètre $\frac{N-i+1}{N}$, de moyenne $\frac{N}{N-i+1}$. Donc,

$$\begin{aligned} E(X_N) &= \sum_{i=1}^{N} E(T_i) \\ &= N\sum_{i=1}^{N} \frac{1}{N-i+1} \\ &= N\sum_{i=1}^{N} \frac{1}{i} \\ &\sim N\log(N) \quad \text{pour N grand} \end{aligned}$$

```
function [p]=vignette(N)
p=0
nontrouves=N
// On introduit un compteur avec le nombre de mathématiciens non trouvés.
while nontrouves>0
   p=p+1;
// Tant qu'il reste encore des mathématiciens à trouver, on reprend une vignett$
   X=grand(1,1,"uin",1,N);
// On suppose que les mathématiciens pas encore trouvés sont ceux numérotés de
// 1 à "non trouvés"; au pire on renumérote. Du coup, si on tire un nombre >
// à non trouvés, c'est qu'on avait déjà ce mathématicien. Dans le cas
// contraire, le nombre de non trouvés descend de 1.
   if X<=nontrouves then nontrouves=nontrouves-1;
   end
end
// La boucle "while" finit quand nontrouves=0: tous les mathématiciens ont été
// trouvés.
endfunction
```